

## THE FOURTH DATA RELEASE OF THE SLOAN DIGITAL SKY SURVEY

JENNIFER K. ADELMAN-McCARTHY,<sup>1</sup> MARCEL A. AGÜEROS,<sup>2</sup> SAHAR S. ALLAM,<sup>1,3</sup> KURT S. J. ANDERSON,<sup>4,5</sup> SCOTT F. ANDERSON,<sup>2</sup>  
 JAMES ANNIS,<sup>1</sup> NETA A. BAHCALL,<sup>6</sup> IVAN K. BALDRY,<sup>7</sup> J. C. BARENTINE,<sup>4</sup> ANDREAS BERLIND,<sup>8</sup> MARIANGELA BERNARDI,<sup>9</sup>  
 MICHAEL R. BLANTON,<sup>8</sup> WILLIAM N. BOROSKI,<sup>1</sup> HOWARD J. BREWINGTON,<sup>4</sup> JARLE BRINCHMANN,<sup>10</sup> J. BRINKMANN,<sup>4</sup> ROBERT J. BRUNNER,<sup>11</sup>  
 TAMÁS BUDAVÁRI,<sup>7</sup> LARRY N. CAREY,<sup>2</sup> MICHAEL A. CARR,<sup>6</sup> FRANCISCO J. CASTANDER,<sup>12</sup> A. J. CONNOLLY,<sup>13</sup> ISTVÁN CSABAI,<sup>7,14</sup>  
 PAUL C. CZARAPATA,<sup>1</sup> JULIANNE J. DALCANTON,<sup>2</sup> MAMORU DOI,<sup>15</sup> FENG DONG,<sup>6</sup> DANIEL J. EISENSTEIN,<sup>16</sup> MICHAEL L. EVANS,<sup>2</sup> XIAOHUI FAN,<sup>16</sup>  
 DOUGLAS P. FINKBEINER,<sup>6</sup> SCOTT D. FRIEDMAN,<sup>17</sup> JOSHUA A. FRIEMAN,<sup>1,18,19</sup> MASATAKA FUKUGITA,<sup>20</sup> BRUCE GILLESPIE,<sup>4</sup> KARL GLAZEBROOK,<sup>7</sup>  
 JIM GRAY,<sup>21</sup> EVA K. GREBEL,<sup>22</sup> JAMES E. GUNN,<sup>6</sup> VIJAY K. GURBANI,<sup>1,23</sup> ERNST DE HAAS,<sup>6</sup> PATRICK B. HALL,<sup>24</sup> FREDERICK H. HARRIS,<sup>25</sup>  
 MICHAEL HARVANEK,<sup>4</sup> SUZANNE L. HAWLEY,<sup>2</sup> JEFFREY HAYES,<sup>26</sup> JOHN S. HENDRY,<sup>1</sup> GREGORY S. HENNESSY,<sup>27</sup> ROBERT B. HINDSLEY,<sup>28</sup>  
 CHRISTOPHER M. HIRATA,<sup>29</sup> CRAIG J. HOGAN,<sup>2</sup> DAVID W. HOGG,<sup>8</sup> DONALD J. HOLMGREN,<sup>1</sup> JON A. HOLTZMAN,<sup>5</sup> SHIN-ICHI ICHIKAWA,<sup>30</sup>  
 ŽELJKO IVEZIĆ,<sup>2</sup> SEBASTIAN JESTER,<sup>1</sup> DAVID E. JOHNSTON,<sup>6</sup> ANDERS M. JORGENSEN,<sup>31</sup> MARIO JURIC,<sup>6</sup> STEPHEN M. KENT,<sup>1</sup> S. J. KLEINMAN,<sup>4</sup>  
 G. R. KNAPP,<sup>6</sup> ALEXEI YU. KNIAZEV,<sup>32</sup> RICHARD G. KRON,<sup>1,18</sup> JUREK KRZESINSKI,<sup>4,33</sup> NIKOLAY KUROPATKIN,<sup>1</sup> DONALD Q. LAMB,<sup>18,34</sup>  
 HUBERT LAMPEITL,<sup>1</sup> BRIAN C. LEE,<sup>35</sup> R. FRENCH LEGER,<sup>1</sup> HUAN LIN,<sup>1</sup> DANIEL C. LONG,<sup>4</sup> JON LOVEDAY,<sup>36</sup> ROBERT H. LUPTON,<sup>6</sup>  
 BRUCE MARGON,<sup>17</sup> DAVID MARTÍNEZ-DELGADO,<sup>37</sup> RACHEL MANDELBAUM,<sup>29</sup> TAKAHIKO MATSUBARA,<sup>38</sup> PEREGRINE M. McGEHEE,<sup>39</sup>  
 TIMOTHY A. MCKAY,<sup>40</sup> AVERY MEIKSIN,<sup>41</sup> JEFFREY A. MUNN,<sup>25</sup> REIKO NAKAJIMA,<sup>9</sup> THOMAS NASH,<sup>1</sup> ERIC H. NEILSEN, JR.,<sup>1</sup>  
 HEIDI JO NEWBERG,<sup>42</sup> PETER R. NEWMAN,<sup>4</sup> ROBERT C. NICHOL,<sup>43</sup> TOM NICHINSKI,<sup>1,44</sup> MARIA NIETO-SANTISTEBAN,<sup>7</sup> ATSUKO NITTA,<sup>4</sup>  
 WILLIAM O'MULLANE,<sup>7</sup> SADANORI OKAMURA,<sup>45</sup> RUSSELL OWEN,<sup>2</sup> NIKHIL PADMANABHAN,<sup>29</sup> GEORGE PAULS,<sup>6</sup> JOHN PEOPLES, JR.,<sup>1</sup>  
 JEFFREY R. PIER,<sup>25</sup> ADRIAN C. POPE,<sup>7</sup> DIMITRI POURBAIX,<sup>6,46</sup> THOMAS R. QUINN,<sup>2</sup> GORDON T. RICHARDS,<sup>6</sup>  
 MICHAEL W. RICHMOND,<sup>47</sup> CONSTANCE M. ROCKOSI,<sup>48</sup> DAVID J. SCHLEGEL,<sup>35</sup> DONALD P. SCHNEIDER,<sup>49</sup>  
 JOSHUA SCHROEDER,<sup>6,50</sup> RYAN SCRANTON,<sup>13</sup> UROŠ SELJAK,<sup>29,51</sup> ERIN SHELTON,<sup>18,19</sup> KAZU SHIMASAKU,<sup>45</sup>  
 J. ALLYN SMITH,<sup>3,31</sup> VERNESA SMOLČIĆ,<sup>52</sup> STEPHANIE A. SNEDDEN,<sup>4</sup> CHRIS STOUGHTON,<sup>1</sup>  
 MICHAEL A. STRAUSS,<sup>6</sup> MARK SUBBARAO,<sup>18,53</sup> ALEXANDER S. SZALAY,<sup>7</sup> ISTVÁN SZAPUDI,<sup>54</sup>  
 PAULA SZKODY,<sup>2</sup> MAX TEGMARK,<sup>55</sup> ANIRUDDHA R. THAKAR,<sup>7</sup> DOUGLAS L. TUCKER,<sup>1</sup>  
 ALAN UOMOTO,<sup>7,56</sup> DANIEL E. VANDEN BERK,<sup>49</sup> JAN VANDENBERG,<sup>7</sup>  
 MICHAEL S. VOGEELEY,<sup>57</sup> WOLFGANG VOGES,<sup>58</sup> NICOLE P. VOGT,<sup>5</sup>  
 LUCIANNE M. WALKOWICZ,<sup>2</sup> DAVID H. WEINBERG,<sup>59</sup> ANDREW A. WEST,<sup>2</sup>  
 SIMON D. M. WHITE,<sup>60</sup> YONGZHONG XU,<sup>61</sup> BRIAN YANNY,<sup>1</sup>  
 D. R. YOCUM,<sup>1</sup> DONALD G. YORK,<sup>18,34</sup> IDIT ZEHAVI,<sup>16</sup>  
 STEFANO ZIBETTI,<sup>58</sup> AND DANIEL B. ZUCKER<sup>32</sup>

*Received 2005 July 29; accepted 2005 September 8*

<sup>1</sup> Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510.

<sup>2</sup> Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195.

<sup>3</sup> Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071.

<sup>4</sup> Apache Point Observatory, P.O. Box 59, Sunspot, NM 88349.

<sup>5</sup> Department of Astronomy, MSC 4500, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003.

<sup>6</sup> Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544.

<sup>7</sup> Center for Astrophysical Sciences, Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218.

<sup>8</sup> Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003.

<sup>9</sup> Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104.

<sup>10</sup> Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal.

<sup>11</sup> Department of Astronomy, University of Illinois, 1002 West Green Street, Urbana, IL 61801.

<sup>12</sup> Institut d'Estudis Espacials de Catalunya/CSIC, Gran Capitá 2-4, E-08034 Barcelona, Spain.

<sup>13</sup> Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260.

<sup>14</sup> Department of Physics of Complex Systems, Eötvös Loránd University, Pf. 32, H-1518 Budapest, Hungary.

<sup>15</sup> Institute of Astronomy and Research Center for the Early Universe, School of Science, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan.

<sup>16</sup> Steward Observatory, 933 North Cherry Avenue, Tucson, AZ 85721.

<sup>17</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218.

<sup>18</sup> Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

<sup>19</sup> Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

<sup>20</sup> Institute for Cosmic Ray Research, University of Tokyo, 5-1-5 Kashiwa, Kashiwa City, Chiba 277-8582, Japan.

<sup>21</sup> Microsoft Research, 455 Market Street, Suite 1690, San Francisco, CA 94105.

<sup>22</sup> Astronomical Institute of the University of Basel, Department of Physics and Astronomy, Venusstrasse 7, CH-4102 Basel, Switzerland.

<sup>23</sup> Lucent Technologies, 2000 Lucent Lane, Naperville, IL 60566.

<sup>24</sup> Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.

<sup>25</sup> US Naval Observatory, Flagstaff Station, 10391 West Naval Observatory Road, Flagstaff, AZ 86001-8521.

<sup>26</sup> Institute for Astrophysics and Computational Sciences and Physics Department, Catholic University of America, Washington, DC 20064.

<sup>27</sup> US Naval Observatory, 3540 Massachusetts Avenue NW, Washington, DC 20392.

<sup>28</sup> Code 7215, Remote Sensing Division, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20392.

<sup>29</sup> Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544.

<sup>30</sup> National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan.

<sup>31</sup> ISR-4, MS D448, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545.

<sup>32</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany.

<sup>33</sup> Obserwatorium Astronomiczne na Suhorze, Akademia Pedagogiczna w Krakowie, ulica Podchorążych 2, PL-30-084 Kraków, Poland.

<sup>34</sup> Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637.

<sup>35</sup> Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720-8160.

## ABSTRACT

This paper describes the Fourth Data Release of the Sloan Digital Sky Survey (SDSS), including all survey-quality data taken through 2004 June. The data release includes five-band photometric data for 180 million objects selected over 6670 deg<sup>2</sup> and 673,280 spectra of galaxies, quasars, and stars selected from 4783 deg<sup>2</sup> of those imaging data using the standard SDSS target selection algorithms. These numbers represent a roughly 27% increment over those of the Third Data Release; all the data from previous data releases are included in the present release. The Fourth Data Release also includes an additional 131,840 spectra of objects selected using a variety of alternative algorithms, to address scientific issues ranging from the kinematics of stars in the Milky Way thick disk to populations of faint galaxies and quasars.

*Subject headings:* atlases — catalogs — surveys

## 1. INTRODUCTION

The Sloan Digital Sky Survey (SDSS) is an imaging and spectroscopic survey of the sky (York et al. 2000) using a dedicated wide-field 2.5 m telescope (Gunn et al. 2005) at Apache Point Observatory, New Mexico. Imaging is carried out in drift-scan mode using a 142 megapixel camera (Gunn et al. 1998) that gathers data in five broad bands, *ugriz*, spanning the range from 3000 to 10,000 Å (Fukugita et al. 1996), on moonless photometric (Hogg et al. 2001) nights of good seeing. The images are processed using specialized software (Lupton et al. 2001; Lupton 2005; Stoughton et al. 2002) and are astrometrically (Pier et al. 2003) and photometrically (Tucker et al. 2005) calibrated using observations of a set of primary standard stars (Smith et al. 2002) on a neighboring 20 inch (0.5 m) telescope. The photometric calibration is accurate to roughly 2% rms in the *g*, *r*, and *i* bands and 3% in *u* and *z*, as determined by the constancy of stellar population colors (Ivezić et al. 2004a; Blanton et al. 2005a), while the astrometric calibration precision is better than 0".1 rms per coordinate (Pier et al. 2003). The median seeing of the imaging data is 1".4 in the *r* band, and the 95% completeness limit in the *r* band is 22.2 mag. All magnitudes are roughly on an AB system (Abazajian et al. 2004) and use the “asinh” scale described by Lupton et al. (1999).

Objects are selected from the imaging data for spectroscopy using a variety of algorithms, including a complete sample of galaxies with reddening-corrected (Schlegel et al. 1998, hereafter SFD98) Petrosian (1976) *r* magnitudes brighter than 17.77

(Strauss et al. 2002), a deeper sample of color- and magnitude-selected galaxies targeting luminous red galaxies (LRGs) from redshift 0.15 to beyond 0.5 (Eisenstein et al. 2001), a color-selected sample of quasars with  $0 < z < 5.5$  (Richards et al. 2002), optical counterparts to *ROSAT* X-ray sources (Anderson et al. 2003), and a variety of stellar and calibrating objects (Stoughton et al. 2002). These targets are arranged on a series of tiles of radius 1°.49 (Blanton et al. 2003), each containing 640 targets and calibration objects, and holes are drilled in the corresponding positions in aluminum plates. Optical fibers with diameter 3" at the focal plane are plugged into these holes and feed the light of the targeted objects to a pair of double spectrographs. The finite size of the fiber cladding means that only one of two objects closer than 55" can be targeted on a given plate; this restriction results in a roughly 10% incompleteness in galaxy spectroscopy (see the discussion in § 3.3). The resulting spectra cover the wavelength range 3800–9200 Å with a resolution of  $\lambda/\Delta\lambda \approx 2000$ . The spectra are flux-calibrated using F subdwarfs, with a broadband uncertainty of 5% (Abazajian et al. 2004); the wavelength calibration uncertainty is roughly 0.05 Å. More than 98% of all spectra are of high enough quality to yield an unambiguous classification and redshift; the remainder are mostly objects fainter than 20 mag and hot featureless stars (e.g., Collinge et al. 2005).

The SDSS began formal operations in 2000 April. Commissioning data taken before that time were released in an Early Data Release (EDR; Stoughton et al. 2002), and subsequent data have been released in roughly yearly intervals (the first, second, and third data releases, DR1, DR2, and DR3; Abazajian et al. 2003, 2004, 2005). This paper describes the fourth data release,

<sup>36</sup> Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

<sup>37</sup> Instituto de Astrofísica de Andalucía (CSIC), Camino Bajo de Huétor, 24 18008 Granada, Spain.

<sup>38</sup> Department of Physics and Astrophysics, Nagoya University, Chikusa, Nagoya 464-8602, Japan.

<sup>39</sup> LANSCE-8, MS H820, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545.

<sup>40</sup> Department of Physics, University of Michigan, 500 East University Avenue, Ann Arbor, MI 48109.

<sup>41</sup> Institute for Astronomy, Royal Observatory, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK.

<sup>42</sup> Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180.

<sup>43</sup> Institute of Cosmology and Gravitation, Mercantile House, Hampshire Terrace, University of Portsmouth, Portsmouth PO1 2EG, UK.

<sup>44</sup> CMC Electronics Aurora, 84 North Dugan Road, Sugar Grove, IL 60554.

<sup>45</sup> Department of Astronomy and Research Center for the Early Universe, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan.

<sup>46</sup> FNRS Institut d’Astronomie et d’Astrophysique, Université Libre de Bruxelles, CP. 226, Boulevard du Triomphe, B-1050 Brussels, Belgium.

<sup>47</sup> Department of Physics, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623-5603.

<sup>48</sup> University of California Observatories/Lick Observatory, University of California, 373 Interdisciplinary Sciences, Santa Cruz, CA 95064.

<sup>49</sup> Department of Astronomy and Astrophysics, 525 Davey Laboratory, Pennsylvania State University, University Park, PA 16802.

<sup>50</sup> Center for Astrophysics and Space Astronomy, University of Colorado, 389 UCB, Boulder, CO 80309.

<sup>51</sup> International Centre for Theoretical Physics, Strada Costiera 11, I-34014 Trieste, Italy.

<sup>52</sup> Department of Physics, University of Zagreb, Bijenička cesta 32, 10000 Zagreb, Croatia.

<sup>53</sup> Adler Planetarium and Astronomy Museum, 1300 Lake Shore Drive, Chicago, IL 60605.

<sup>54</sup> Institute for Astronomy, 2680 Woodlawn Road, Honolulu, HI 96822.

<sup>55</sup> Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

<sup>56</sup> Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101.

<sup>57</sup> Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104.

<sup>58</sup> Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85741 Garching, Germany.

<sup>59</sup> Department of Astronomy, Ohio State University, 140 West 8th Avenue, Columbus, OH 43210.

<sup>60</sup> Max-Planck-Institut für Astrophysik, Postfach 1, D-85748 Garching, Germany.

<sup>61</sup> Theoretical Division, MS B285, Los Alamos National Laboratory, Los Alamos, NM 87545.

hereafter DR4, consisting of survey data meeting our quality specifications taken through 2004 June. Data releases are cumulative; DR4 includes all data released in previous data releases. There have been no substantive changes to the imaging or spectroscopic software since DR2; thus, DR4 includes the identical data to DR2 and DR3.

Finkbeiner et al. (2004) describe the release of SDSS imaging data taken outside the formal SDSS footprint (as described by York et al. 2000); most of these data are at low Galactic latitudes. In addition, there have been a number of value-added catalogs of SDSS data, including galaxies (Blanton et al. 2005a), quasars (Schneider et al. 2005), clusters of galaxies (Miller et al. 2005), compact groups (Lee et al. 2004), merging galaxies (Allam et al. 2004), white dwarf stars (Kleinman et al. 2004), and asteroids (Ivezić et al. 2004b).

In § 2 of this paper we describe the fourth data release itself. This release includes spectra from 209 plates principally in the southern Galactic cap that explore a variety of extensions of the main spectroscopic targeting algorithms; these are described in § 3. Additional new features of DR4 can be found in § 4. In § 5 we describe a recently discovered subtle effect in the SDSS data, whereby misestimated sky levels near bright galaxies systematically bias the counts of neighboring faint galaxies. We conclude in § 6.

## 2. WHAT IS INCLUDED IN DR4

As is described by York et al. (2000), the SDSS imaging data are taken along a series of stripes, great circles on the sky that aim to fill a contiguous area in the northern Galactic cap, and three noncontiguous stripes in the southern Galactic cap. The top panel of Figure 1 shows the region of sky included in DR4. As this figure shows, the northern Galactic cap is currently covered by two contiguous regions, one centered roughly on the celestial equator, and the other at around  $\delta = +40^\circ$ . New data taken since DR3 (indicated in gray in the figure) have reduced the size of the gap between these regions; these new data lie on the northern edge of the equatorial patch or on the southern edge of the northern patch. The images cover  $6670 \text{ deg}^2$ . The great circle stripes overlap at the poles of the survey; roughly 25% of the region is covered more than once; the repeat data in the northern Galactic cap are available in DR4 as well.

These data are processed by the same version of the imaging pipelines used in DR2 and DR3. There have been concerns in previous data releases that the SDSS imaging flat fields change with time, especially in  $u$ . We have confirmed that the flat fields used for the data taken since DR3 introduce errors of at most 2% in photometry in the  $u$  chips; they are accurate to roughly 1% for all other chips.

The bottom panel of Figure 1 shows the sky coverage of the spectroscopic data. As the spectroscopy necessarily lags the imaging, it covers less area, a total of  $4783 \text{ deg}^2$ . Calculation of this area is more accurate than in previous data releases (indeed Abazajian et al. [2003] erroneously indicated the spectroscopic area of DR3 to be  $4188 \text{ deg}^2$ ; the correct value for DR3 is  $3732 \text{ deg}^2$ ). The SDSS Catalog Archive Server (CAS) database includes a package that computes the spherical polygons corresponding to all geometric objects in the survey, defined by the boundaries of the imaging runs and the spectroscopic plates (Gray et al. 2004; see also the discussion by Blanton et al. 2005a).<sup>62</sup> Given plate overlaps, one can divide up the spectroscopically tiled sky into

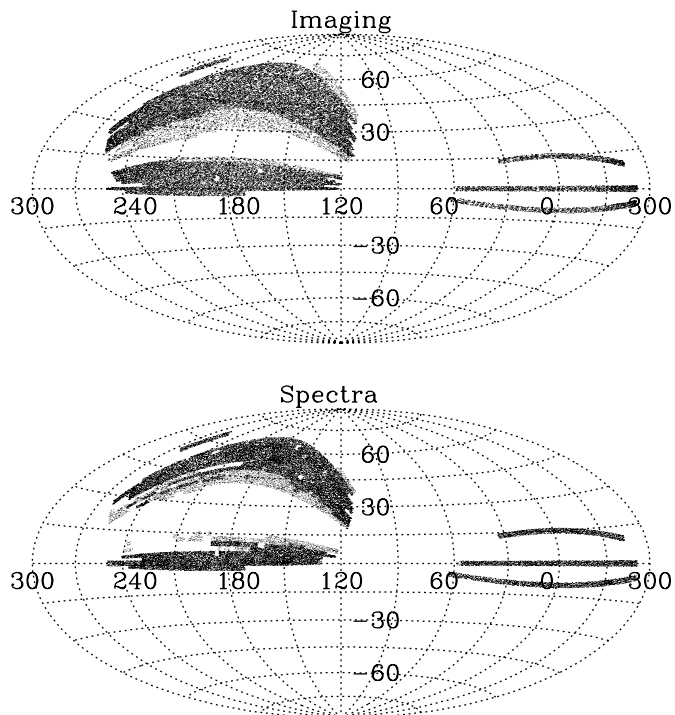


FIG. 1.—Distribution on the sky of SDSS imaging (*top*) and spectroscopy (*bottom*) included in DR4, shown in J2000.0 equatorial coordinates. These cover  $6670$  and  $4783 \text{ deg}^2$ , respectively. The regions of sky that are new to DR4 are shaded more lightly, corresponding to 26% and 28% increments, respectively, over DR3.

nonoverlapping “sectors,” each one of which defines a region uniformly covered by the same tiles. Summing the area of those sectors that cover an observed spectroscopic plate gives the total area of the spectroscopic sample.

The spectroscopic data include 673,280 spectra, arrayed on 1052 plates of 640 fibers each. Thirty-two fibers per plate are devoted to measurements of sky. There are approximately 480,000 galaxies, 64,000 quasars, and 89,000 stars among the spectra; only 1% of the spectra are unclassified. About 2% of the objects are repeated on adjacent plates, as a check of the reproducibility of the spectroscopy.

A number of spectroscopic plates have been observed more than once; this data release includes those duplicate observations that satisfy our signal-to-noise ratio (S/N) criteria. These criteria are a mean S/N per pixel of 4 for objects with  $g_{\text{fiber}} = 20.2 \text{ mag}$  in the blue cameras of the spectrographs, and similarly for  $i_{\text{fiber}} = 19.9 \text{ mag}$  in the red cameras. In particular, there are 52 plates with more than one observation that are released. The majority of these plates have only one duplicate observation, but three plates (406, 419, and 483) have two duplicates, and two plates (406 and 412) have four. Note that in most cases, the plates were replugged before being reobserved, so the correspondence between fiber number and object will be different on each observation. These duplicate plates are useful for monitoring the robustness of specific spectroscopic features, especially in low-S/N spectra (e.g., Hall et al. 2004), increasing the S/N of spectra by co-adding, and checking for variability (Wilhite et al. 2005; although note that this paper used many spectroscopic observations that fall below our S/N cut and are therefore not included in the release).

In the autumn months, when the southern Galactic cap is visible in the Northern Hemisphere, the SDSS imaging has been confined to a stripe along the celestial equator, plus two “outrigger”

<sup>62</sup> This package is not yet publicly available with DR4 but will be included in a future data release.

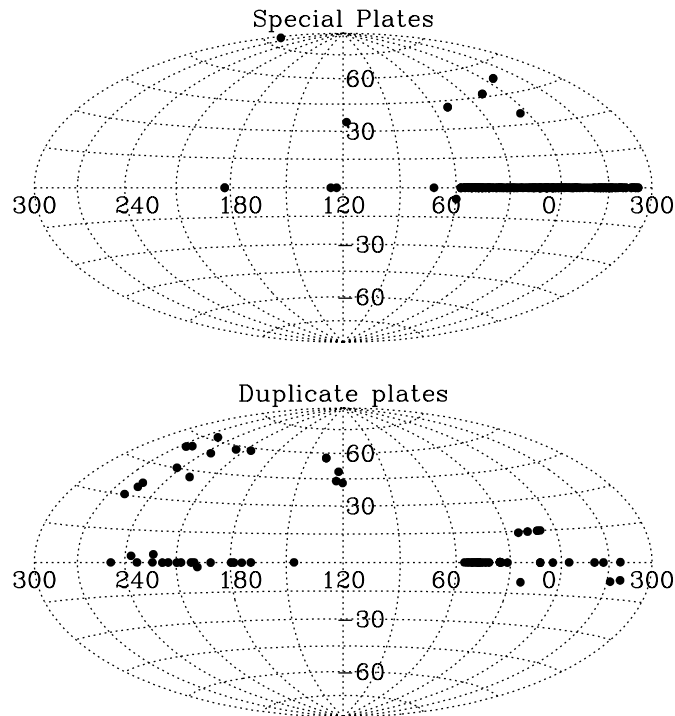


FIG. 2.—Distribution on the sky of special plates (*top*) and duplicate plates (*bottom*) included in DR4, shown in J2000.0 equatorial coordinates. The majority of the special plates are on the celestial equator, in the southern equatorial cap, but there are several exceptions, as this figure makes clear.

stripes, centered roughly at  $\delta = +15^\circ$  and  $-10^\circ$ , respectively (these are visible on the right-hand side of the panels of Fig. 1). We have performed multiple imaging passes of the equatorial stripe, data that are being used for a deep co-addition, as well as a variety of variability studies (e.g., Ivezić et al. 2003). These multiple imaging scans are planned for inclusion in a future release.

Given the relatively small footprint of the imaging in the southern Galactic cap, we finished the spectroscopy of targets selected by our normal algorithms quite early in the survey; most of these data were included already in DR1. We carry out imaging only under pristine conditions: the Moon is below the horizon, the sky is cloudless, and the seeing is good. To make optimal use of the remaining time, we undertook a series of spectroscopic observing programs, based mostly on the imaging data of the equatorial stripe in the southern Galactic cap, designed to go beyond the science goals of the main survey. DR4 includes 206 plates, totaling 131,840 spectra, from these programs, carried out in the autumn months of 2001, 2002, and 2003. These programs are described in the next section. In addition, DR4 includes nine plates that are repeat observations of these “southern” plates. The sky distributions of both the special and repeat plates are given in Figure 2.

As described by Stoughton et al. (2002), these data are available both as flat files (the Data Archive Server) and via a flexible Web interface to the SDSS database (the CAS).<sup>63</sup> A summary of the characteristics of DR4 may be found in Table 1.

### 3. THE SPECIAL SOUTHERN SPECTROSCOPIC PLATES

Unlike the spectroscopic plates for the main survey, the special southern plates did not aim to produce a complete, uniform

TABLE 1  
CHARACTERISTICS OF THE SDSS FOURTH DATA RELEASE (DR4)

Parameter	Value
<b>Imaging</b>	
Footprint area (deg <sup>2</sup> ).....	6670 (26% increment over DR3)
Imaging catalog.....	180 million unique objects
Magnitude limits: <sup>a</sup>	
<i>u</i> (mag).....	22.0
<i>g</i> (mag).....	22.2
<i>r</i> (mag).....	22.2
<i>i</i> (mag).....	21.3
<i>z</i> (mag).....	20.5
Median PSF width (arcsec).....	1.4 in <i>r</i>
Astrometry.....	<0".1 rms absolute per coordinate
<b>Spectroscopy, Main Survey</b>	
Footprint area (deg <sup>2</sup> ).....	4783 (28% increment over DR3)
Wavelength coverage (Å).....	3800–9200
Resolution $\lambda/\Delta\lambda$ .....	1800–2100
Number of plates.....	1052
Number of spectra.....	673,280
<b>Additional Spectroscopy</b>	
Number of special plates.....	206
Number of spectra on special plates.....	131,840
Number of plates, repeat observations.....	70

<sup>a</sup> 95% completeness for point sources in typical seeing; 50% completeness numbers are typically 0.4 mag fainter (DR1 paper).

survey of targets over the entire survey area. Indeed, many of the programs carried out with these plates were quite exploratory in nature, designed to probe the limits of the main survey target algorithms, push deeper in various categories of targets, or select classes of objects not fully explored by the main algorithms (especially stars). These selection algorithms changed several times in the course of the 2001–2003 time period, which makes keeping track of the target selection flags somewhat complicated. Table 2 shows which algorithms (as described below) are used on which spectroscopic plates. This table makes reference to a *chunk*, i.e., a region of imaging data on which target selection was run and spectroscopic tiles defined (Blanton et al. 2003). The different chunks are numbered but are also given names (e.g., lowz97), which can be useful in queries of the CAS. While some chunks include only a few plates, others include several tens of plates. Among the most important of these are the various merged programs. Chunk 22, the first of the major southern programs, used simple extensions of the main survey targeting algorithms (although the plates included some objects targeted with the original algorithms as well). Chunks 48 and 73 combined a variety of algorithms to target a range of different targets. Chunks 45, 52, 62, 74, and 97 used double-length exposures to target faint galaxies and stars.

As described by Stoughton et al. (2002), the survey marks each object targeted for spectroscopy with flag bits encoded in the *primTarget* and *secTarget* quantities. The relevant flag hex bits for the southern plates are given in Table 3. With a few exceptions, these special plates are flagged with the highest bit in *secTarget*. The *primTarget* flags unfortunately can take on different meanings depending on chunk. Thus, using these target flags requires taking the chunk number or plate numbers into account. Table 3 also indicates the total number of objects with each of the target flags.

<sup>63</sup> See <http://www.sdss.org/dr4>.

TABLE 2  
SPECIAL CHUNKS AND PLATES

Chunk	Chunk Name	Plates	Program	Comments
97.....	lowz97	1859	low- <i>z</i> galaxies, BCG, deep LRG	<sup>a</sup>
83, 96.....	seguetest83, seguestest96	1662–1664, 1857	SEGUE test	
84.....	seguetest84	1665–1666	Perseus	<sup>b</sup>
79.....	photoz79	1629, 1632, 1633, 1635	Photo- <i>z</i>	
74.....	lowz74	1538, 1539, 1561–1566	Low- <i>z</i> galaxies, BCG, deep LRG	<sup>a</sup>
73.....	merged73	1473–1476, 1485–1499, 1504–1506, 1508–1511, 1514–1518, 1521–1523, 1529	Complete main, <i>u</i> -band galaxies, variability, high proper motion stars, all point sources, fainter LRG	<sup>c</sup>
72.....	fstar72	1468, 1471, 1472	M31: F stars, quasars	<sup>b</sup>
62.....	lowz62	1243	Low- <i>z</i> galaxies, deep LRG	<sup>a</sup>
52.....	lowz52	1156, 1157	Low- <i>z</i> galaxies, deep LRG	<sup>a</sup>
51.....	fstar51	1149, 1150, 1152–1155	F stars	<sup>a</sup>
50.....	rockosi50	1133–1135, 1137, 1143	MS turnoff stars	
49.....	mun49	1118–1132	Thick/thin disk	
48.....	merged48	1062–1096, 1101, 1103–1107, 1114–1117	Complete main, <i>u</i> -band galaxies, double-lobed radio sources, faint quasars, all point sources	<sup>c</sup>
45.....	lowz45	1021–1024, 1026–1037	Low- <i>z</i> galaxies, deep LRG	<sup>a</sup>
29.....	photoz29	807–810	Photo- <i>z</i>	
	annis29	802–806	Low- <i>z</i> galaxies	
	loveday29	811	Low- <i>z</i> galaxies	
	fstar29	797	F stars	
22.....	south22	673–714	Main extension: galaxies, quasars	<sup>d</sup>
21.....	photoz21	669–672	Photo- <i>z</i>	
	Schlegel/Locus	323–324	Stellar locus	

<sup>a</sup> Double-length exposures.

<sup>b</sup> Imaging not available.

<sup>c</sup> Standard merged program.

<sup>d</sup> Extensions of standard algorithms.

All of the programs described below put a bright limit of 15 mag in *g*, *r*, and/or *i* fiber magnitudes (i.e., aperture magnitudes of 3'' diameter, corresponding to the entrance aperture of the spectroscopic fibers), to prevent saturation and excessive cross talk in the spectrographs. Unless otherwise indicated, the magnitudes used below are point-spread function (PSF) magnitudes for stellar and quasar work and are Petrosian (1976) magnitudes for galaxy targets. Similarly, for galaxies, colors are based on fits to exponential or de Vaucouleurs profiles convolved with the PSF (Stoughton et al. 2002; Abazajian et al. 2004; we refer to these as model colors in what follows). All magnitudes and colors have been corrected for SFD98 Galactic extinction. With a few exceptions, objects for which spectroscopy already existed from previous spectroscopic observations were excluded as targets. This means that defining complete samples requires combining objects from more than one selection criterion. In all equations that follow, *u*, *g*, *r*, *i*, and *z* are expressed in magnitudes.

### 3.1. Galactic Kinematic Programs

Many of the southern plate selection algorithms were designed to select stars of various types, mostly for kinematic studies of the Galactic halo and disks.

*F stars.*—F stars are numerous in the Galaxy, have sharp spectral features allowing accurate radial velocities to be measured, are approximate standard candles if the stars are on the main sequence, and are of high enough luminosity that they can be seen to great distances. This program aims to use F star radial velocities to understand the kinematics of the outer parts of the Milky Way. Most of these plates are in the southern Galactic cap equatorial stripe (chunks fstar29 and fstar51), but there are also three plates in the vicinity of M31 (fstar72) and another two plates

centered roughly on the Perseus cluster (seguetest84). The M31 (e.g., Zucker et al. 2004a, 2004b) and Perseus imaging data are *not* included in DR4 but will be included in a future data release. A program to target quasars in the plates near M31 is described in § 3.6; similarly, see § 3.3 for a program to target galaxies in the Perseus plates.

For the F star program, stellar objects with  $-0.3 < (g - r) < 0.3$  and with  $19.0 < g < 20.5$  were selected. Given the faint magnitudes of these stars and the desire for accurate radial velocities, these plates were observed to an S/N per pixel of 5 for stars with  $g = 20.2$ , requiring roughly twice the usual spectroscopic exposure of 45 minutes.

*Main-sequence turnoff.*—This program was designed to study the kinematics and metallicities of high-latitude thick-disk and halo stars. Stars were selected in two SDSS spectroscopic tiles, one centered at  $(l, b) = (64^\circ, -45^\circ)$  and the other at  $(l, b) = (114^\circ, -62^\circ)$ . In both tiles, stars were selected to be  $g - r < 0.8$  and  $r < 19.15$ . The tile at  $(l, b) = 114^\circ, -62^\circ$  is shared with the thick/thin-disk program described below, and there is an additional cut at  $i - z > 0.2$  to avoid overlap. The data from this program that appear in DR4 are a randomly selected subset of the total number of available targets in both tiles. The exposure time for this program was the standard spectroscopic exposure time for the main survey.

*Thick/thin-disk stars.*—A third program was focused on the kinematics of the thin and thick disks, targeting a complete sample of bright ( $i < 18.26$ ), red ( $i - z > 0.2$ ), late-type stars in three adjacent spectroscopic tiles centered at high Galactic latitude ( $l = 123^\circ, b = -63^\circ$ ). It provides an in situ sample of thin- and thick-disk stars with radial velocities, proper motions, and spectroscopic metallicity determinations, densely

TABLE 3  
TARGET SELECTION FLAGS

Chunk	Target	primTarget	Number	Comments
84.....	Perseus: galaxies	0x80000040	787	a
	Perseus: F stars	0x80002000	313	a
83, 96.....	SEGUE test	...	2116	a,b
74, 97.....	BCG	0x80000100	439	a
73.....	Complete main	...	2020	c
	<i>u</i> -band-selected galaxies: priority	0x80000040	765	a
	<i>u</i> -band-selected galaxies: extra	0x800000C0	2128	a
	<i>u</i> -band-selected galaxies: extra2	0x80000140	2108	a
	Variability: high priority	0x81000000	318	a
	Variability: low priority	0x80800000	660	a
	High proper motion stars	0x80010000	966	a
	All point sources: $L > 1$	0x80000001	3934	a
	All point sources: $L < 1$	0x80000002	5478	a
	Faint LRG	0x840000A0	2301	a
72.....	M31: F stars	0x80002000	1623	a
	M31: quasars	0x80000004	144	a
51.....	F star	0x80002000	3531	a
50.....	MS turnoff	0x80002000	2947	a
49.....	Thick/thin disk	0x80040000	8859	a
48.....	Complete main	...	1605	c
	<i>u</i> -band-selected galaxies	0x80000040	3426	a
	Double radio sources/bent	0x80200000	97	a
	Double radio sources/straight	0x80000010	13	a
	Faint southern quasars	...	6655	a,b
	All point sources	0xA0000000	14779	a
45, 52, 62, 74, 97.....	Low- <i>z</i>	0x80000040	14210	a
	Deep LRG/cut I	0x80000020	1230	a
	Deep LRG/cut II	0x84000020	205	a
29.....	Low- <i>z</i> (Loveday)	0x80000040	570	a
	Low- <i>z</i> (Annis)	0x80000040	2849	a
	F star	0	583	d
22.....	Main	...	8008	c
	Main extension	...	16760	b,e
21, 29, 79.....	Photo- <i>z</i>	0x80000040	4526	a
	Stellar locus	0x00002000	1188	d

<sup>a</sup> The southern plates are flagged with the highest bit set in secTarget; i.e., secTarget = 0x80000000.

<sup>b</sup> Regular primTarget flags, with the highest bit set.

<sup>c</sup> Regular primTarget and secTarget flags.

<sup>d</sup> Stars of interest are flagged STAR\_BHB in the Objtype\_name column of specObjAll in the CAS.

<sup>e</sup> Regular secTarget flags, with the highest bit set.

sampling a single line of sight out to 2 kpc above the Galactic plane.

*SEGUE test plates.*—SEGUE (Sloan Extension for Galactic Understanding and Exploration; Newberg et al. 2003) is part of a follow-up project to the SDSS, emphasizing spectra of stars to study stellar populations and Galactic structure. The targets on the SEGUE test plates in DR4 are Galactic stars of a variety of colors and magnitudes, meant to sample the range of available spectral types and to test the reproducibility of radial velocity measurements at faint magnitudes. SEGUE aims to probe the structure of the Milky Way by sampling horizontal branch, F turnoff, and G dwarf and K dwarf and giant stars, representing a variety of distances in the disk and halo. These plates are being used to refine target selection for the SEGUE program itself and will be described in detail in a future paper.

### 3.2. Other Stellar Target Selection Algorithms

*All point sources.*—As part of an exploration of the full stellar locus, as well as a search for unusual objects of all sorts, we carried out a survey of all point sources. These objects were assigned to available spectroscopic fibers on the so-called merged

program plates, which included a mixture of mostly extragalactic targets. In 2002 (chunk 48), this included a random sampling of all point sources with clean photometry (see the discussion of fatal and nonfatal flags by Richards et al. 2002) with SFD98-corrected *i*-band PSF magnitudes brighter than 19.1. Not surprisingly, the vast majority of the targets were chosen from the densest core of the stellar locus in color-color space.

In chunk 73, we revised the algorithm to put greater weight on the wings of the stellar locus. We defined a distance from the ridge of the stellar locus, by asking for the median and standard deviation  $u - g$ ,  $g - r$ , and  $i - z$  of stars in narrow bins of  $r - i$ . Having tabulated these, we calculated a crude  $\chi^2$ -like quantity:

$$L = \frac{1}{3} \sum_{u-g, g-r, i-z} \left( \frac{\text{color} - \text{median color}}{\text{standard deviation}} \right)^2. \quad (1)$$

A total of 75% of all stars had  $L < 1$ . We sparsely sampled the objects with  $L < 1$ , decreasing the sampling rate for smaller  $L$ . As Table 3 describes,  $L > 1$  and  $L < 1$  objects are given different flags.

These spectra were used for a determination of the completeness of the quasar target selection algorithm (Vanden Berk et al.

2005). The overwhelming majority of these objects are confirmed to be stars; only 10 of the 19,543 of the all-point-sources targets analyzed in that paper are quasars not previously targeted as such.

*High proper motion.*—Munn et al. (2004) matched USNO-B astrometric data (Monet et al. 2003) with SDSS to determine proper motions of stars (§ 4). We targeted for spectroscopy stars with high proper motion along the equatorial stripe in the southern Galactic cap. The stars selected had  $r < 19.5$ , and two cuts were made on proper motion. The first was a simple cut of proper motion  $\mu > 100 \text{ mas yr}^{-1}$ . The second used a cut in reduced proper motion  $H_r$ , as follows. Define  $H_r$  as

$$H_r = r + 5 \log_{10} \mu + 5. \quad (2)$$

The sample was defined by

$$\begin{aligned} g - i &< 2.000, & H_r &> 16, \\ 2.000 < g - i &< 2.375, & H_r &> 8 + 4(g - i), \\ 2.375 < g - i, & & H_r &> 17.5. \end{aligned} \quad (3)$$

This cut removes the majority of main-sequence red dwarfs, which would otherwise swamp the sample. This sample is useful for studying the kinematics and spectral properties of very nearby objects, especially white dwarfs and objects near the end of the main sequence. Despite the two cuts to define the sample, these are marked with a single target selection flag.

*Stellar locus.*—For plates 323 and 324, stellar targets were chosen from SDSS imaging data, randomly sampling the stellar locus in color space, in order to explore the full range of stellar spectra. In particular, grids of width 0.04 mag in the  $(u - g)$ ,  $(r - i)$  and  $(r - i)$ ,  $(i - z)$  color planes were set down, and a single star at each grid point was chosen, if the grid had any stars at all. This process was repeated until each plate had 600 targets assigned. These objects have the target flag STAR\_BHB set.

### 3.3. Galaxy Target Selection Algorithms: Low Redshifts

*Main extension: galaxies.*—In chunk 22, observed in autumn of 2001, we used direct extensions of the main survey target selection algorithms. In particular, the main galaxy sample (Strauss et al. 2002) was modified only slightly, by removing the cut on objects with half-light Petrosian  $r$ -band surface brightness  $\mu_{50,r}$  below 24.5 mag arcsec $^{-2}$ . This adds less than one object per square degree.

For LRGs, the sample was changed to probe galaxies fainter and bluer than those of the main survey selection described by Eisenstein et al. (2001). The cuts that were used garner about 40 objects deg $^{-2}$ :

$$\begin{aligned} r &< 19.5, \\ r &< 13.4 + c_{\parallel}/0.3, \\ \mu_{50,r} &< 25, \\ |c_{\perp}| &< 0.4 \end{aligned} \quad (4)$$

for cut I and

$$\begin{aligned} r &< 19.5, \\ g - r &> 1.65 - c_{\perp}, \\ \mu_{50,r} &< 25, \\ r_{\text{PSF}} - r_{\text{model}} &> 0.3 \end{aligned} \quad (5)$$

for cut II. Here  $c_{\parallel} \equiv 0.7(g - r) + 1.2(r - i - 0.18)$  and  $c_{\perp} \equiv (r - i) - (g - r)/8.0$ . In particular, equation (5) pushes the cut I selection 0.3 mag fainter, while equation (6) significantly loosens the primary color cut of cut II, to more fully explore the transition between cut I and cut II.

Note that these plates also include an extension to the quasar target selection algorithm (§ 3.6).

*Complete main galaxies.*—The main galaxy sample, as described by Strauss et al. (2002), is as complete as we can make it. However, we obtain spectra for only about 90% of these objects. This incompleteness has several causes, including the fact that two spectroscopic fibers cannot be placed closer than 55'' on a given plate, possible gaps between the plates, fibers that fall out of their holes, and so on. This program aims to observe the remaining 10% of galaxy targets in order to have a region of sky with truly complete galaxy spectroscopic coverage. This is particularly important for studies of galaxy pairs, which are by definition strongly affected by the 55'' rule. The target selection algorithm is simple: all galaxies selected by the algorithm described by Strauss et al. (2002) in the southern Equatorial stripe, minus those galaxies that actually have a successfully measured redshift in routine targeting. A complete sample of quasars and LRGs was similarly defined and targeted on the same plates.

*The  $u$ -band galaxies.*—The main galaxy target selection is carried out in the  $r$  band and is flux limited at  $r = 17.77$ . The bluest galaxies have  $u - r \approx 0.6$ ; thus, the sample is complete for galaxies to  $u \approx 18.4$ . In order to further explore the  $u$ -band luminosity density and recent history of star formation in the universe, we carried out a deeper  $u$ -band survey of galaxies in the SDSS. The sample consists of galaxies whose SFD98-corrected magnitudes satisfy the following criteria:  $u_{\text{select}} < 19.8$ ,  $g < 20.5$ , and  $17.5 < r < 20.5$ , which had not been previously spectroscopically targeted. Here the  $u$ -band magnitude is calculated from the  $r$ -band Petrosian magnitude and model colors:  $u_{\text{select}} = r_{\text{Petro}} + (u - r)_{\text{model}}$ , as this quantity has better noise properties than does  $u_{\text{Petro}}$ . Like main galaxy target selection, a cut was placed on  $r$ -band half-light surface brightness,  $\mu_{50,r} < 24.5$ , to exclude extremely low surface brightness galaxies. Similarly, a cut on  $r_{\text{PSF}} - r_{\text{model}} > 0.2$  excludes stars. Baldry et al. (2005) describe the sample in detail and present the resulting  $u$ -band luminosity function.

There were also two categories of lower priority  $u$ -band-selected targets for chunk 73. The “extra”  $u$ -band targets relaxed the selection criteria described in the previous paragraph slightly:

$$\begin{aligned} u_{\text{select}} &< 20.0 \text{ or } u_{\text{model}} < 19.8 \text{ or } u_{\text{Petro}} < 19.5, \\ g &< 20.7; & 17.5 < r < 20.7, \\ \mu_{50,r} &< 24.7; & r_{\text{PSF}} - r_{\text{model}} > 0.15, \end{aligned} \quad (6)$$

while the “extra2” targets relaxed the magnitude limit further:

$$\begin{aligned} u_{\text{select}} &< 20.3, \\ g &< 20.5; & 17.5 < r < 20.5, \\ \mu_{50,r} &< 24.5; & r_{\text{PSF}} - r_{\text{model}} > 0.20. \end{aligned} \quad (7)$$

Note that the objects targeted in this program do not include objects with spectroscopic observations already in hand. Thus, one needs to combine objects from various programs to define a complete sample. See Baldry et al. (2005) for determination of the completeness of the resulting sample.

*Low-redshift galaxies.*—For studies of the faint end of the galaxy luminosity function (e.g., Blanton et al. 2005b), the small-scale clustering of galaxies, and galaxy populations in clusters, it

TABLE 4  
SELECTION FRACTIONS FOR LOW- $z$  GALAXY ALGORITHM

Magnitudes	Redshift	Chunk 45	Chunk 52	Chunk 62	Chunk 74	Chunk 97
$17.77 \leq r < 19.0$ .....	$0.00 \leq z_p < 0.15$	1.0	0.85	0.7	1.0	1.0
	$0.15 \leq z_p < 0.20$	0.15	0.1275	0.105	0.15	0.3
	$0.20 \leq z_p < 0.25$	0.15	0.1275	0.105	0.15	0.3
$19.0 \leq r < 19.5$ .....	$0.00 \leq z_p < 0.15$	0.25	0.2	0.15	0.25	0.3
	$0.15 \leq z_p < 0.20$	0.25	0.16	0.1	0.25	0.3
	$0.20 \leq z_p < 0.25$	0.25	0.18	0.15	0.25	0.3
$17.77 \leq r < 19.0$ .....	$0.25 \leq z_p$	0.015	0.015	0.015	0.17	0.3
	$z_p < 0.0$	0.65	0.65	0.65	0	0
$19.0 \leq r < 19.5$ .....	$0.25 \leq z_p$	0.005	0.005	0.005	0.15	0.3
	$z_p < 0.00$	1.0	1.0	1.0	0	0

is useful to have a survey that probes to as low luminosities as possible. With the apparent magnitude limit of the main galaxy survey, the volume to which galaxies with absolute magnitudes fainter than, say,  $M_r = -17$  are probed is quite small. With this in mind, we carried out a survey of low-redshift objects, selected to nearly 2 mag fainter than the SDSS main sample limit of  $r = 17.77$ . Our redshift selection used photometric redshifts derived from second-order polynomial fits to observed Petrosian  $r$  magnitudes and model colors, with separate fits done in bins of model  $g - r$  color. For chunks 45, 52, and 62 we used the SDSS EDR photometry and spectroscopy then available to derive photometric redshifts, while for chunks 74 and 97 we were able to derive improved photometric redshifts using catalog-co-added stripe 82 SDSS photometry, combined with all available SDSS redshift data on the southern equatorial stripe as of 2003 July 11. This included much of the data taken for the express purpose of calibrating the photometric redshift relation in the SDSS photometric system (see § 3.4).

Galaxies were then chosen for observations based on their photometric redshift  $z_p$  and Petrosian magnitude  $r$ . In particular, the aim was to target as complete a sample as possible for  $17.77 \leq r < 19.0$  and true redshift below 0.15, and sparse samples to higher redshifts, as well as at fainter magnitudes  $19.0 \leq r < 19.5$ . The specific target categories, in order of highest to lowest priority for fiber assignment, are listed in Table 4. The sparse sampling fraction values were chosen to get reasonable distributions of objects over the target categories and to keep approximately similar target distributions from chunk to chunk, which resulted in somewhat different sampling fractions for each of the five chunks in which this algorithm was used, as given in Table 4. The tabulated sampling fractions give the fraction of targets made available for fiber assignment; the actual fraction of targets with spectra is lower, in particular for the lowest priority targets. A caveat to note is that chunk 45 used the star/galaxy separation criteria of the SDSS photometric pipeline to select galaxies, and this resulted in noticeable contamination of stars in several of the chunk 45 plates located at lower galactic latitudes. The other chunks used the star/galaxy cut employed by the SDSS main sample target selection algorithm ( $r_{\text{PSF}} - r_{\text{model}} \geq 0.3$  or 0.24, depending on the version of the photometric pipeline), which is more conservative for selecting galaxies.

Two additional southern programs, both in chunk 29, also selected samples of low-redshift and/or low-luminosity galaxies. One plate, 811 (chunk loveday29), was also devoted to low-luminosity galaxies but used a different algorithm. The sample was chosen as all objects from the EDR photometric redshift catalog (Csabai et al. 2003) with  $z_p > 0.003$ ,  $r_{\text{Petro}} < 20$ , and estimated absolute magnitude  $M_r > -18$ .

Finally, plates 802–806 in chunk 29 (annis29) used a somewhat simpler photometric redshift relation and selected those objects with  $i_{\text{Petro}} \leq 20$ ,  $i_{\text{Petro}} + (r - i)_{\text{model}} \geq 17.75$ , and  $z_p \leq 0.17$ – $0.19$ . The photometric redshift limit on each plate was chosen to give enough targets to match the available number of fibers.

*Perseus*.—Imaging scans were taken centered roughly on the Perseus cluster at  $z \approx 0.018$  and were used to spectroscopically target galaxies. The double exposure plates 1665, 1666 targeted, and obtained redshifts for, approximately 400 galaxies as faint as  $r_{\text{fiber}} = 18.8$  in a region centered on the cluster at  $(\alpha, \delta) = (49^\circ 96', 41^\circ 53')$ . The majority of the galaxies are associated with the cluster, although there are 50 objects in a background overdensity at  $z \approx 0.05$  (Brunzendorf & Meusinger 1999). These plates also included approximately 300 foreground F stars, as described in § 3.1.

#### 3.4. Galaxy Target Selection Algorithms for Calibrating Photometric Redshifts

*Photo- $z$* .—The SDSS five-band photometry goes substantially fainter than does the spectroscopy, suggesting the opportunity to derive photometric redshifts for vastly more objects than have spectroscopy (see, e.g., Csabai et al. 2003). Calibrating the photometric redshift relation requires a training sample exploring the same range of apparent magnitudes and colors as the objects for which photometric redshifts will eventually be derived. The SDSS LRG sample (Eisenstein et al. 2001) obtains spectra for red faint ( $r < 19.5$ ) galaxies; photometric redshifts of this relatively uniform population (e.g., Eisenstein et al. 2003) are fairly robust (e.g., Padmanabhan et al. 2005). However, we do not have a corresponding sample of faint blue galaxies for the calibration of photometric redshifts. Therefore, a series of spectroscopic plates was designed to obtain redshifts for the blue end of the galaxy color distribution at the faint end.

Galaxies were selected using the following cuts, designed to select objects with photometric redshift greater than 0.3 (Csabai et al. 2003):

$$\begin{aligned}
 &0.40 + 0.6(u - g) < g - r < 1.7 - 0.1(u - g), \\
 &-0.5 < u - g < 3.0; \quad 0 < g - r < 1.8, \\
 &-0.5 < r - i < 1.5; \quad -1 < i - z < 1.5, \\
 &18.0 < u < 24.0; \quad 18.0 < g < 21.5, \\
 &17.8 < r < 19.5; \quad 16.5 < i < 20.5, \\
 &16.0 < z < 20.0; \quad \sigma_u < 0.6; \quad \sigma_{g,r,i,z} < 0.25. \quad (8)
 \end{aligned}$$

For objects that satisfied the above cuts, the quantity  $\exp(c\{(g - r) - [0.40 + 0.6(u - g)]\})$  was calculated; if it was larger than a



random number chosen between 0 and 1, the object was targeted for spectroscopy. The coefficient  $c = 0.1411$  was chosen to obtain an appropriate density of targets. Note that plates 672 and 809 have the same center, and some objects were inadvertently observed twice.

### 3.5. Selection of Higher Redshift Galaxies

*Deep LRG exposures.*—The SDSS LRG sample (Eisenstein et al. 2001) targets high-redshift ( $0.2 < z < 0.55$ ), luminous galaxies by their colors and magnitudes. As part of southern targeting, we have extended this algorithm in several ways. The first, the deep LRG sample, uses double-length spectroscopic exposures to get higher S/N spectra of LRGs with  $z > 0.25$ . Combining with previous spectroscopy therefore gives a total exposure time 3 times the standard value. These observations serve two purposes: first, to obtain measurements of velocity dispersion for galaxies where the current, single-pass spectroscopy is only “good enough for a redshift;” and second, given the discontinuity in targeting algorithm at  $z \approx 0.4$  (the distinction between cut I and cut II; see Eisenstein et al. 2001), higher S/N spectra allow the exploration of possible spectroscopic changes due to differences in stellar populations at this transition.

*Faint LRGs.*—The LRG cut II sample aims for a flux-limited sample of LRGs with redshifts roughly between 0.40 and 0.55. We also experimented with an extension of this cut, going substantially fainter. The faint LRG sample has the following criteria:

$$\begin{aligned} 17.5 < i_{\text{dev}} < 20, \\ i_{\text{Petro}} < 19.1, \\ 0.5 < g - r < 3.0, \\ 0.0 < r - i < 2.0, \\ c_{\parallel} \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6, \\ d_{\perp} \equiv (r - i) - (g - r)/8.0 > 0.5. \end{aligned} \quad (9)$$

Here  $i_{\text{dev}}$  refers to the de Vaucouleurs model magnitude in the  $i$  band, and all colors are model colors. Given the typical  $r - i \approx 0.7$  for LRGs, this is about 0.3 mag deeper than the  $r < 19.5$  cut II sample, limited at  $r = 19.5$ . There are additional cuts, to select against, in sequence, very concentrated objects, very low surface brightness objects, and stellar objects, respectively:

$$\begin{aligned} i_{\text{fiber}} - i_{\text{dev}} > 0, \\ i_{\text{fiber}} < 21.4, \\ i_{\text{PSF}} - i_{\text{model}} > 0.15, \\ r_{\text{PSF}} - r_{\text{model}} > 0.4. \end{aligned} \quad (10)$$

Finally, objects that were targeted as part of normal target selection were removed. These objects were observed on plates that received the standard exposure times.

*BCGs.*—A separate program explicitly targeted the brightest galaxies in clusters. While LRGs are often the brightest galaxies in their clusters (e.g., Loh & Strauss 2006), they need not be. The so-called MaxBCG method described by Bahcall et al. (2003) searches for galaxies with the apparent magnitudes and colors of LRGs, together with a red sequence of fainter elliptical galaxies in the vicinity (e.g., Gladders & Yee 2000). The BCG program targeted BCG candidates found with this method, with estimated redshifts in the range  $0.4 < z < 0.7$ . The MaxBCG algorithm was run on photometry derived from co-adding the detections (at

the catalog level) of multiple scans of the southern equatorial stripe.

### 3.6. Other Target Selection Algorithms

*Main extension: quasars.*—In chunk 22, the quasar target selection algorithm (Richards et al. 2002) was extended as follows. For objects selected from the *ugri* color cube, the magnitude limit was changed from  $i = 19.1$  to 19.9, while for the *griz* color cube (where high-redshift quasars are selected), the limit changed from  $i = 20.2$  to 20.4.

As Richards et al. (2002) describe, there are regions outside the stellar locus that are heavily contaminated by hot white dwarfs, M dwarf–white dwarf pairs (Smolčić et al. 2004), and other non-quasars. These regions are explicitly excluded from quasar target selection in the main survey; however, in the extension, they are allowed back in. Similarly, there is a region of color space where  $z \approx 2.7$  quasars intersect the stellar locus. In the main survey, objects falling in this region are sparse sampled to 10% (to reduce the number of stars); in the extension, all objects falling in the mid- $z$  box defined in Richards et al. (2002) are targeted.

*Faint quasars.*—The quasar target selection was further modified for the faint quasar targets on the merged plates on chunk 48. In particular, the following are the changes over the standard quasar target selection algorithm described in Richards et al. (2002):

1. The magnitude limit is set to  $i = 20.1$ , rather than  $i = 19.1$  for the *ugri* color cube.
2. The magnitude limit for optical counterparts to FIRST (Becker et al. 1995) sources is set to  $i = 20.65$ , rather than  $i = 19.1$ .
3. The standard quasar target selection algorithm requires that the estimated PSF magnitude errors in  $u$  and  $g$  both be less than 0.1 for UV excess sources ( $u - g < 0.6$ ). This limit is now set to 0.2.
4. For UV excess objects with  $19.1 \leq i < 20.1$ , we add a requirement that  $g - r < 0.7$  to minimize stellar contamination.
5. Objects in the mid- $z$  box are excluded altogether.
6. In addition to the usual “distance from the stellar locus” algorithm used to target quasars, in the main sample there are hard color cuts used to select high-redshift quasars in the *griz* color cube. These color cuts are not used in the southern targeting.
7. Finally, there were additional color cuts to reject unphysical objects affected by bad CCD columns; we required  $g - r > -0.5$ ,  $r - i > -0.5$ , and  $i - z > -0.6$ .

*M31 quasars.*—Low-redshift quasars were targeted on the M31 imaging data (chunk 72) using the standard quasar selection algorithm (Richards et al. 2002), but excluding the high-redshift candidates selected from the *griz* cube. The confirmed quasars can be used to probe gas in the halo of M31. The plates used for this program also included F star targets, as described in § 3.1.

*Double-lobed radio sources.*—Quasar target selection selects unresolved optical counterparts to FIRST radio sources, while “serendipity” target selection (Stoughton et al. 2002) selects FIRST sources with extended optical counterparts. This works fine for compact or core-dominated radio sources but is less effective for double-lobed radio sources, in which the optical counterpart is associated with a point between the two lobes. An algorithm was developed to find these double-lobed sources, including allowing for the more challenging case of *bent* double sources. In particular, pairs of FIRST sources separated by  $90''$  or less without SDSS optical counterparts were identified. Given the distance  $d$  between the centers of these two sources, a rectangle is drawn centered at the midpoint between the sources, with dimensions  $0.57d$ ,  $1.33d$ , with the short axis parallel to the line

connecting the two sources. This box size was chosen empirically to include the core of a sample of bent double sources compiled by E. Blanton (see Blanton et al. 2001). Optical counterparts with  $r < 19.8$  that fell into the box were selected as bent double counterparts. A subset of those objects, which fell into a  $12''$  square centered on the midpoint between the sources, were selected as straight double counterparts.

*Variability.*—As discussed above, the southern equatorial stripe has been imaged multiple times in the course of the SDSS, allowing photometric variability to be studied. Variable sources were selected for follow-up spectroscopy using pairs of observations of unique unsaturated point sources with  $i < 21$  and requiring that the changes in the  $g$  and  $r$  bands exceed 0.1 mag and are at least  $3\sigma$  significant (using error estimates computed by the photometric pipeline). The time difference between the two observations varied from 56 to 1212 days.

This target selection produced targets on the southern equatorial stripe with a surface density of  $\sim 10$  objects  $\text{deg}^{-2}$ , after removing those objects that already had SDSS spectra. The majority of these targets have colors consistent with  $z < 2$  (UV excess) quasars and RR Lyrae stars, classifications confirmed by the spectroscopy. Two classes of targets were assigned, based on apparent magnitude: the high-priority objects were those with  $i$  magnitude brighter than 19.5, and the low-priority objects had  $19.5 < i < 21$ . In practice, all targets were tiled on the plates observed.

#### 4. OTHER NEW FEATURES OF DR4

The DR3 paper (Abazajian et al. 2005) described a quality flag calculated for each field in the SDSS imaging data. This flag was based on the measured PSF, the difference between aperture and PSF magnitudes of stars (a diagnostic of errors in the determination of the PSF), and the location of the stellar locus in color-color space.<sup>64</sup> With DR4, we make available the detailed plots on which these quantities are based. Much of the rationale of the quantities plotted is given by Ivezić et al. (2004a). For each of the 204 imaging runs included in DR4, a Web page is available with detailed statistics and plots of the quantities used in the determination of the quality, as well as tests of the uniformity of the sky levels from one camera column to another and statistics on the relative astrometry between the different photometric bands. These statistics are summarized in a table<sup>65</sup> giving the fraction of the fields in each run of a given quality; these then allow an overall determination of the quality of each run. A total of 85% of the 204 runs show median offsets in the  $s$  principal color between runs of less than 0.02 mag in all six camera columns.

The DR2 paper (Abazajian et al. 2004) describes the availability of proper motions for each object based on matching with USNO-B (Monet et al. 2003). The DR4 CAS also includes the recalibrated proper motions based on using the SDSS astrometry to improve the USNO-B positions (Munn et al. 2004).

<sup>64</sup> Abazajian et al. (2005) described a bug whereby the quality was based on only the  $s$  color (Helmi et al. 2003), as opposed to all four available principal colors that describe the stellar locus. Unfortunately, this bug has not yet been fixed as of this writing.

<sup>65</sup> Available at [http://das.sdss.org/dr4/data/imaging/QA/summaryQA\\_analyzePC.html](http://das.sdss.org/dr4/data/imaging/QA/summaryQA_analyzePC.html).

#### 5. OBJECT COUNTS NEAR BRIGHT OBJECTS

A study of weak lensing in the SDSS imaging data (Mandelbaum et al. 2005) found a systematic, 5% decrease in the number density of faint objects within  $90''$  of bright ( $r < 18$ ) galaxies. This is due to a systematic overestimation of the sky levels in the vicinity of bright objects, which of course will affect all the measured photometric quantities of faint objects, including the classification as stars and galaxies. The brighter the object, the larger the overestimation of sky; indeed, restricting to foreground galaxies brighter than  $r = 16$ , the number density of fainter galaxies is 10% below the mean. This is due to the way in which the sky levels are estimated in the SDSS; they can be biased upward by the faint outer isophotes of a bright galaxy (Stoughton et al. 2002; see also the discussion by Strauss et al. 2002). Closer than  $40''$  to the bright objects, the intrinsic clustering of galaxies makes it difficult to assess the problem. This problem is most relevant for applications that involve correlating positions of bright objects with faint ones, yielding a spurious anticorrelation on the affected scales. We are currently investigating changes to the imaging pipeline to address this problem.

#### 6. CONCLUSIONS

We have presented the SDSS Fourth Data Release, consisting of five-band imaging data over  $6670 \text{ deg}^2$  and spectra for over 800,000 objects. These data represent roughly a 27% increment over the previous data release (DR3; Abazajian et al. 2005). The data releases are cumulative, and the DR4 includes all data of DR2 and DR3, reduced with the same version of the imaging and spectroscopic pipelines.

The next data release is planned for mid-2006 and will consist of all SDSS survey-quality data gathered through 2005 June. The immediate goals of the SDSS are to fill the gap between the two portions of the northern Galactic hemisphere seen in Figure 1 in both imaging and spectroscopy. This, together with new surveys to study structure in the Milky Way (Newberg et al. 2003) and supernovae at  $0.1 < z < 0.3$  (Frieman et al. 2004), will keep the SDSS facilities busy through summer 2008.

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max-Planck Society. The SDSS Web site is <http://www.sdss.org>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

#### REFERENCES

- Abazajian, K., et al. 2003, *AJ*, 126, 2081
- . 2004, *AJ*, 128, 502
- . 2005, *AJ*, 129, 1755
- Allam, S. S., Tucker, D. L., Smith, J. A., Lee, B. C., Annis, J., Lin, H., Karachentsev, I. D., & Laubscher, B. E. 2004, *AJ*, 127, 1883
- Anderson, S., et al. 2003, *AJ*, 126, 2209
- Bahcall, N., et al. 2003, *ApJS*, 148, 243
- Baldry, I. K., et al. 2005, *MNRAS*, 358, 441
- Becker, R. H., White, R. L., & Helfand, D. J. 1995, *ApJ*, 450, 559
- Blanton, E. L., Gregg, M. D., Helfand, D. J., Becker, R. H., & Leighly, K. M. 2001, *AJ*, 121, 2915
- Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, *AJ*, 125, 2276
- Blanton, M. R., et al. 2005a, *AJ*, 129, 2562

- Blanton, M. R., et al. 2005b, *ApJ*, 631, 208
- Brunzendorf, J., & Meusinger, H. 1999, *A&AS*, 139, 141
- Collinge, M., et al. 2005, *AJ*, 129, 2542
- Csabai, I., et al. 2003, *AJ*, 125, 580
- Eisenstein, D. J., et al. 2001, *AJ*, 122, 2267
- . 2003, *ApJ*, 585, 694
- Finkbeiner, D. P., et al. 2004, *AJ*, 128, 2577
- Frieman, J., et al. 2004, *BAAS*, 205, 120.01
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Gladders, M. D., & Yee, H. K. C. 2000, *AJ*, 120, 2148
- Gray, J., Szalay, A. S., Fekete, G., O'Mullane, W., Nieto-Santesteban, M. A., Thakar, A. R., Heber, G., & Rots, A. H. 2004, *There Goes the Neighborhood: Relational Algebra for Spatial Data Search*, Microsoft Tech. Rep. MSR-TR-2004-32
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- . 2005, *AJ*, submitted
- Hall, P. B., Snedden, S. A., Niederste-Ostholt, M., Eisenstein, D. J., Strauss, M. A., York, D. G., & Schneider, D. P. 2004, *AJ*, 128, 534
- Helmi, A., et al. 2003, *ApJ*, 586, 195
- Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, *AJ*, 122, 2129
- Ivezić, Ž., et al. 2003, *Mem. Soc. Astron. Italiana*, 74, 978
- . 2004a, *Astron. Nachr.*, 325, 583
- . 2004b, *SDSS Moving Object Catalog, Data Release 1*, NASA Planetary Data System, 9, 1
- Kleinman, S. J., et al. 2004, *ApJ*, 607, 426
- Lee, B. C., et al. 2004, *AJ*, 127, 1811
- Loh, Y. S., & Strauss, M. A. 2006, *MNRAS*, in press (astro-ph/0510500)
- Lupton, R. H. 2005, *AJ*, submitted
- Lupton, R. H., Gunn, J. E., Ivezić, Ž., Knapp, G. R., Kent, S., & Yasuda, N. 2001, in *ASP Conf. Ser. 238, Astronomical Data Analysis Software and Systems X*, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 269
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, *AJ*, 118, 1406
- Mandelbaum, R., et al. 2005, *MNRAS*, 361, 1287
- Miller, C., et al. 2005, *AJ*, 130, 968
- Monet, D. G., et al. 2003, *AJ*, 125, 984
- Munn, J. A., et al. 2004, *AJ*, 127, 3034
- Newberg, H. J., et al. 2003, *BAAS*, 203, 112.11
- Padmanabhan, N., et al. 2005, *MNRAS*, 359, 237
- Petrosian, V. 1976, *ApJ*, 209, L1
- Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, *AJ*, 125, 1559
- Richards, G. T., et al. 2002, *AJ*, 123, 2945
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525 (SFD98)
- Schneider, D. P., et al. 2005, *AJ*, 130, 367
- Smith, J. A., et al. 2002, *AJ*, 123, 2121
- Smolčić, V., et al. 2004, *ApJ*, 615, L141
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Strauss, M. A., et al. 2002, *AJ*, 124, 1810
- Tucker, D., et al. 2005, *AJ*, submitted
- Vanden Berk, D. E., et al. 2005, *AJ*, 129, 2047
- Wilhite, B. C., Vanden Berk, D. E., Kron, R. G., Schneider, D. P., Pereyra, N., Brunner, R. J., Richards, G. T., & Brinkmann, J. V. 2005, *ApJ*, 633, 638
- York, D. G., et al. 2000, *AJ*, 120, 1579
- Zucker, D., et al. 2004a, *ApJ*, 612, L117
- . 2004b, *ApJ*, 612, L121